# REPORT

# Evaluation of SNAP Employment and Training Pilots

## OMB Supporting Statement
## Part B: Collections of Information Employing Statistical Methods

## February 11, 2019

# CONTENTS

# EXHIBITS

# ATTACHMENTS

A       Agriculture Act of 2014 (The 2014 Farm Bill)

B       Study Description: Research Questions, Data Sources, and Key Outcomes

C.1     Registration Document – English

C.2     Registration Document – Spanish

C.3     Registration Document – Screenshots

**ATTACHMENTS** (*continued*)

**ATTACHMENTS** (*continued*)

**ATTACHMENTS** (*continued*)

B.1. Description of respondent universe and sampling methods

**Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.**

The Evaluation of the SNAP E&T Pilots study will assess the effectiveness of pilot studies in ten Grant awardee sites: California Department of Social Services, Fresno Bridge Academy; State of Delaware Department of Health and Human Services; Georgia Division of Family and Children Services; Illinois Department of Human Services; Kansas Department for Children and Families; Commonwealth of Kentucky Cabinet for Health and Family Services, Department for Community Based Services; Mississippi Department of Human Services; Vermont Agency of Human Services; Virginia Department of Social Services; and Washington State Department of Social and Health Services. Separate study designs were developed and implemented for each site. A summary of key features of each of the pilot projects can be found in Attachment DD. All sites use designs where individuals are the units of random assignment. Some sites implemented single-arm treatment designs, where individuals are randomly assigned into either a treatment group that receives a set of services or to a control group. Other sites will implement multi-arm treatment designs, where individuals are randomly assigned to one of several treatment groups or a control group. The remaining sites implemented multi-arm treatment designs with enhancements. For these designs, some services in one treatment group are an "add-on" to the

services offered in another treatment group. For example, one treatment might consist of job search assistance (JSA) only while another treatment would consist of JSA and support services. The target populations (universes) are different across the sites. Some sites are targeting larger populations of SNAP work registrants or ABAWDs, while other sites have more narrowly focused target populations.

This section describes the sampling procedures used for the pilot studies, including: (1) sample frame determination, (2) design features specific to individual-level RA design studies, (3) sampling plan, and (4) expected response rates and nonresponse analysis.

## 1. Sample Frame Determination

Each pilot study design dictates how the evaluation team identifies the sample frame for that pilot. Our initial interactions with pilots focused on providing them information about the study and also obtaining information about how the site will implement both the intervention and its evaluation. In communicating with the sites, the evaluation team focused on the following issues:

- Where components of the study intake process—obtaining consent (Attachments D.1, D.2, D.3, and D.4), collecting baseline information via a registration document (Attachments C.1 and C.2), conducting RA, and service provider assessments—take place in the overall program flow of the pilot

- How random assignment (RA) will be integrated into program intake

- What enhancements might be appropriate in the intervention definition to ensure a strong treatment–counterfactual distinction

- What geographic locations within the site the study should include

- What adjustments, if any, are needed and possible in the participant-tracking management information system (MIS) to ensure receipt of reliable data on pilot services

- How best and most securely to transfer data to our contractor

- How agency and provider staff will guard against participants receiving services that are inconsistent with their RA status

- What partner and provider organizations will be involved

This information helped the evaluation team define the study population of interest for each site and determine how to construct sample frames for each site.

## 2.  Design Features

Different designs have different implications for the research question being addressed, their statistical power, their feasibility, and the ease with which staff can adhere to evaluation procedures. The evaluation team worked with each site to develop designs best suited for their interventions.

A key design variable across all sites is the unit of RA—whether individuals or clusters of individuals, such as those within the same SNAP office, are randomly assigned to a treatment or control group. All ten pilots proposed randomly assigning individual participants within SNAP offices to receive different interventions, which creates greater statistical power and thus greater ability to detect program impacts, compared with randomly assigning entire offices to offer specific interventions (or no intervention at all).

The evaluation team deployed one of three specific RA models to address the research questions of interest:

- *Single-arm.* Under this design, individuals are randomly assigned into either a treatment group that receives the intervention services or a control group that does not.

- *Multi-arm with distinct treatments.* In multi-arm designs, individuals are randomly assigned to one of several treatment groups that receive distinct intervention services or to a control group that does not receive any of them. For example, one treatment group might receive job search assistance while another receives more intensive case management and the control group receives no E&T services.

- *Multi-arm with enhancements.* In some multi-arm designs, some services in one treatment group can be an "add-on" to the services offered in the other treatment group. For example, one treatment might consist solely of job search assistance, while another would consist of job search assistance plus supportive services. This design provides a measure not only of the impact of each intervention, but also of the value added from the more intensive group-based services relative to individual-level services.

Depending on each Grantee's design, SNAP participants were randomly assigned to one of potentially several treatment groups or to a control group. After SNAP participants provided consent to participate in the study and completed a baseline registration document via EPIS, a

web-based system designed by the evaluation team, Grantees' pilot intake staff submitted the information to the evaluation team with the click of a button. The evaluation team then randomly assigned the SNAP participant and returned the assignment status within seconds of receiving it.

## 3. Sampling plan

**<u>Participants.</u>** Data measuring the evaluation's outcomes will come from both administrative and survey data. The evaluation team will collect administrative data for the primary study outcomes measuring employment status, earnings, and public assistance receipt (Attachment T). Because administrative data will be available for all study participants, the evaluation team will not conduct sampling procedures for collecting this information. Through followup surveys (Attachments O.1 – O.6), however, the evaluation team will obtain more detailed measures of primary outcomes, such as measures of job quality and wages, and secondary outcomes including food security, health and well-being, and housing stability. Followup surveys will be administered to a sample of pilot participants.

The evaluation team will employ a two-phase sampling approach for collecting data on secondary outcomes. Two-phase sampling or double sampling involves a phase 1 sampling from the full population and then, in phase 2, selecting a second sample from the nonrespondents from phase 1. In the second phase, more focused resources (in this case, field follow-up) is devoted to the sample, which leads to higher response rates among the phase 1 nonrespondents than what would happen in the absence of the more focused field follow-up. This sampling approach typically increases response rates, reduces the potential for nonresponse biases, and reduces costs and burden associated with survey data collection, while maintaining a sufficient level of power to detect differences in outcomes between treatment and control groups within pilot sites.

The full pilot study sample will be around 52,800 individuals. The evaluation team plans to randomly sample approximately half of these individuals—around 25,000—for the 12-month

followup survey and plan to collect data from 18,240 individuals using two-phase sampling. During phase 1 of the two-phase sampling design, the evaluation team will attempt to conduct a CATI survey to collect data from the 25,000-person subsample starting approximately in November 2016 and continuing on a rolling basis for sampled participants. The evaluation team expects to obtain responses from about 14,500 individuals (58 percent of the first-phase sample). Among the 10,500 nonresponders, the evaluation team will randomly select 50 percent for field followup data collection (5,250 CATI non-responders). With intensive field follow-up, our goal is to obtain a 71.24 percent response rate or higher from the second phase sample, resulting in total sample size of around 18,240 or more completes, which includes initial CATI respondents and nonrespondents. This approach will lead to an overall weighted response rate over 80 percent, which provides some cushion in case the expected CATI response rate of 58 percent and second phase response rate of 71 percent are difficult to achieve. During the second followup data collection period, the evaluation team will employ the same two-phase sampling approach starting with the 18,240 or so completes who completed the first followup survey. The second followup data collection will start around November 2018 and continue on a rolling basis for sampled participants.

Although the evaluation team expects this two-phase sampling approach to yield a response rate of 80 percent or higher for the followup surveys the evaluation team realizes that the response rate could fall below 80 percent. The evaluation team anticipates a response rate between 65 and 80 percent, but assumes a response rate of 80 percent when discussing estimation procedures and statistical power in B.2.

For the participant focus groups and case study interviews, convenience samples were used to recruit participants as described in A.2.3.d.

**Employers.** For the employer focus groups, convenience samples were also used. Employers participating in the SNAP E&T pilot were targeted based on their role in the pilot projects. To the extent there is variation in the type of employers involved in the pilot, the evaluation team will attempt to recruit employers from various industries. The contractor will work with Grantee and provider staff to identify employers for the focus groups. The state staff will reach out to the employers to gage interest and make introductions. The contractor will follow-up via email or by telephone as needed (Attachment L.3).

## 4. Response Rates and Nonresponse Bias Analysis

In each site, follow-up surveys (Attachments O.1, O.2, O4, and O.5) will be conducted with a randomly selected subsample of treatment and control participants at multiple points in time. An 80 percent response rate for each follow-up survey is anticipated. The evaluation team will implement a number of steps to maximize response rates, including the use of two-phase sampling. More details about this and other steps the study will take are described in section B3. The surveys will collect data on service receipt and outcomes from both treatment and control group members in each site. They will build on other surveys that have been administered successfully with similar low-income populations, such as the surveys used in WIA GSE (OMB Control Number 1205-0482, Discontinued September 30, 2014) and Rural Welfare-to-Work (OMB Control Number 0970-0246, Discontinued January 31, 2007). The 36-month follow-up survey (Attachments O.4 and O.5) will be largely consistent with the earlier follow-up survey with only minor changes such as confirming information collected at the 12-month follow-up. Each follow-up survey will be translated into Spanish by a certified bilingual translator using the Referred Forward Translation approach in which a translator having extensive experience in survey development first translates the questionnaire, and a second translator reviews that work and recommends changes in phrasing or wording, or dialectical variations. The two then meet to

discuss comments and determine the preferred questionnaire wording.  Each follow-up survey will be administered via computer-assisted telephone interview (CATI).

To assess whether nonresponse bias exists, the evaluation team obtained baseline information via a registration document from the study samples at the time of random assignment. Baseline data include name; contact information; demographic characteristics such as age, gender, and education level; household characteristics such as the presence of children and household income; and baseline values of primary outcome measures including employment status, earnings, and participation in SNAP, TANF, and Medicaid. Using the baseline data, the evaluation team will (1) compare survey follow-up respondents and nonrespondents within the treatment and comparison groups, (2) test the significance of differences between respondents' and nonrespondents' characteristics, (3) look at whether these differences are the same across treatment and comparison groups, and (4) compare characteristics of the respondent and nonrespondent samples with those of the frame.

## B.2. Procedures for the collection of information

**Describe the procedures for the collection of information including:**

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**
- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

This section discusses the procedures the evaluation team will use for impact estimation and the anticipated statistical power of the estimates. Data collection activities and procedures for the focus groups, interviews, and surveys are described in Part A of this information collection request.

## 1.  Estimation Procedures

With an experimental design, unbiased impact estimates can be obtained from the differences between mean outcomes in different research groups. By using regression procedures that control for highly predictive covariates, however, the evaluation team will improve the precision of estimates and adjust for small baseline differences between groups that may arise by chance or from survey nonresponse or missing administrative records data.

For the study's primary outcomes that are continuous variables, such as quarterly earnings, the evaluation team will use ordinary least squares regression to estimate impacts. For other primary outcomes that are binary variables, such as those measuring employment and the receipt of SNAP benefits, the evaluation team plans to use a logistic regression model to estimate impacts. For the logistic models, the evaluation team also will test whether the results are sensitive to our modeling approach by estimating impacts on the primary outcome using a linear probability model as an alternative approach.

The evaluation team will use a staged approach to assess the effects of the pilots using analytical models carefully aligned with the specific evaluation designs in each site. First, to examine the *overall* average effects of the interventions, the evaluation team will estimate the following regression model for site s:

$$Y_{is} = \alpha + \delta\, T_{is} + \beta\, X_{is} + \varepsilon_{is}$$

where Y is the outcome of interest (e.g., quarterly earnings) for individual *i* in site *s*, $\alpha$ is the regression intercept, *T* is a binary treatment indicator that equals 1 for individuals in *any* treatment group (e.g., $T_1$ or $T_2$ and *0* for control individuals), *X* represents a set or vector of individual characteristics, $\beta$ is a vector of regression coefficients for those characteristics, and $\varepsilon$ is the regression's residual. The parameter of interest, $\delta$, represents the site's average impact on treatment individuals, and the associated t-statistic can be used to gauge statistical significance.

The covariates most likely to improve precision are baseline levels of the outcome measure. Other covariates in the model will include people's work experience and demographic and economic characteristics.

Second, for sites with multiple treatment arms including service enhancements, the evaluation team will estimate the impact of *each* type of intervention, relative to the overall control group and to each other, using the model:

$$Y_{is} = \alpha + \delta_1 T_{1,is} + \delta_2 T_{2,is} + \beta X_{is} + \varepsilon_{is}.$$

All the variables are the same as in the previous model, except there is a separate binary treatment indicator for each intervention ($T_1$ and $T_2$). The parameters of interest, $\delta_1$, $\delta_2$, and $(\delta_1 - \delta_2)$ represent the site's average impacts on people receiving $T_1$ compared to the control group, $T_2$ compared to the control group, and $T_1$ compared to $T_2$, respectively. This model can be easily adapted to reflect the presence of a suitability screen in which agencies assess people's suitability for specific services by replacing the variables $T_{1,is}$ and $T_{2,is}$ with $T_{1,is} \times I_{1,is}$ and $T_{2,is} \times I_{2,is}$, where $I_{2,is}$ is a binary variable set equal to 1 for individuals suitable to receive $T_1$ services and *0* for individuals not suitable and $I_{2,is}$ is a binary variable set equal to 1 for individuals suitable to receive $T_2$ services and *0* for individuals not suitable. The evaluation team will conduct F-tests to examine the joint statistical significance of $\delta_1$ and $\delta_2$ to assess whether either intervention improved participants' outcomes. If the F-statistic is significant, the evaluation team will examine the separate t-statistics for each intervention.

**Accounting for sample design and survey nonresponse.**  As noted in section B.1.3, FNS's goal is a 65-80 percent response rate for the follow up surveys. If an 80 percent response rate is not achieved on a given survey in a pilot a nonresponse bias analysis will be conducted following OMB standards and guidelines to assess relationships between participant characteristics and

nonresponse. The evaluation team will test for the significance of differences in respondents' and nonrespondents' characteristics, and whether these differences are the same in the treatment and control/comparison groups.

To account for the two-phase sampling design (see section B3 for more details) and survey nonresponse, the evaluation team will construct weights for all analyses using the survey data so that the sample of respondents to each follow-up survey represent those of the baseline samples within each site. A base weight for each survey respondent will be calculated as the inverse of his/her probability of selection for the survey. These base weights will also be adjusted to account for nonresponse. The evaluation team will use propensity scoring, first estimating statistical models to predict the likelihood that a person will respond to the survey and then using that propensity score to construct nonresponse weighting adjustment factors. Variables that are available for both respondents and nonrespondents (from the registration document and administrative data) will be used to create the weighting adjustment factors.

**Analyzing policy-relevant service enhancements to address national issues.** To the extent possible, our plans are to incorporate enhancements to site interventions to learn about key program features that have national implications. The analysis will examine the effects of interventions with common features across sites by grouping sites to conduct a "pooled" site analysis. In these cases, the evaluation team will estimate a similar set of models as above, except that each treatment indicator would represent the common component or characteristic, such as whether there are financial incentives or whether the services were group- or individually-focused.

**Adjusting for no-shows and crossovers.** In any experiment in the real world, some members of the treatment group may not receive intervention services (no-shows), and some

controls may be exposed to the interventions (crossovers). To correct for these sample members, the evaluation team will use an instrumental variable (IV) approach by replacing the $T_1$ and $T_2$ indicators in the models above with the indicator variables $P_1$ and $P_2$ that equal 1 for those who received intervention services and 0 for those who did not, and the evaluation team will use $T_1$ and $T_2$ as instruments for $P_1$ and $P_2$.

**Subgroup analyses.** For each demonstration project, the analysis will estimate impacts among subgroups of participants who may respond differently to the intervention. One component of our subgroup analysis will be to estimate impacts among subgroups of participants, defined by their baseline characteristics, who may respond differently to the intervention(s). In particular, the analysis will estimate impacts on subgroups defined by:

- Family composition (e.g., whether single or married, whether a parent, presence of children in the household, and presence of more than one adult in the household)
- Labor force attachment (e.g., recent employment experiences)
- Baseline earnings (e.g., whether the person had zero or positive earnings before the pilot)
- History of SNAP receipt (e.g., whether the person has participated before the current spell)
- Demographic characteristics (e.g., age, gender, race/ethnicity, education)
- Extent of barriers to employment (e.g., language, lack of transportation)
- Income (e.g., less than 100 percent of poverty level)

The evaluation team will estimate impacts for subgroups identified before analysis by adding to the empirical model a term that interacts the treatment indicator with an indicator for whether the participant is in the subgroup. The coefficient on this term provides an estimate of program impact for the subgroup. Multiple treatment-by-subgroup indicators for subgroups with more than two levels (such as race/ethnicity) will be included and F-tests will be used to assess whether differences in impacts across subgroup levels are statistically significant.

**Using the implementation analysis to inform the impact analysis.** To dig deeper into the impact estimates, the evaluation team will use data from the implementation analysis to

statistically test whether key measurable program features—such as types of intervention services, program organization and partnerships, the target populations, and the extent to which the interventions were implemented as planned—are associated with cross-site variations in impacts. The evaluation team will estimate impacts for program-related subgroups by re-formulating the models from above as the first level of a multilevel model or hierarchical linear model. In the second level, the evaluation team will regress the estimated impacts of each site (estimates of $\delta_s$) on various characteristics of the site obtained from the planned implementation analysis to help identify best practices at the grantee programs.

## 2. Statistical Power

To adequately address the evaluation's research questions, the design must have sufficient statistical power to detect impacts that are policy relevant and of practical significance. The evaluation design will allow detection of policy-relevant impacts in each of the demonstration sites, overall and for key sample subgroups. The sample sizes needed for the study were determined by focusing on minimum detectable impacts (MDIs) for the primary outcomes of employment, earnings, and SNAP receipt. However, the MDIs for other secondary outcomes such as food insecurity to be examined in the evaluation were also considered.

Each pilot study will be sufficiently powered to detect impacts for the primary outcomes and key secondary outcomes for the full sample and key subgroups. The evaluation team presents minimum detectable impacts (MDIs) in Exhibit B.2a for the primary outcome measures and for the secondary outcome of food insecurity for a single pilot site under the assumption of 90 percent confidence and 80 percent power. MDIs were calculated for impacts comparing individuals from both treatment arms to the control condition (T vs. C) and for impacts comparing individuals from a single treatment arm to the control condition ($T_j$ vs. C). The evaluation team presents MDIs for two levels of enrollment of 3,000 participants per pilot and

5,000 participants per pilot, which captures the range of sample sizes for 7 out of 10 pilot projects. (The remaining three pilot projects have larger sample sizes of 5,252 participants, 7,520 participants, and 14,000 participants, which will result in smaller MDIs.)

For the individual-level RA design with a sample size of 5,000, the evaluation team will be able to detect a 2.9-percentage-point increase in the employment rate for the treatment group from the overall pilot (combining $T_1$ and $T_2$ and comparing to C) and a 3.3-percentage-point increase in the employment rate for each intervention, $T_1$ and $T_2$, separately (comparing $T_1$ to C and $T_2$ to C). MDIs for earnings and public assistance participation are also presented in Table B.2a. For a subgroup of half the size of the overall sample—such as people in single-parent SNAP households—the MDIs would be about 4.0 and 4.7 percentage points, respectively. The table shows MDIs for a sample size of 5,000 per site. For smaller sites with 3,000 participants, the MDIs are about 30 percent higher in designs 1 and 2. For example, the MDI for employment would be 3.7 instead of 2.9 percentage points. For the percentage of people who are food secure, with 5,000 participants per site the evaluation team will be able to detect a 6.4-percentage-point increase in the food security rate for the treatment group from the overall pilot combining $T_1$ and $T_2$ and comparing to C) and a 7.4-percentage-point increase in the food security for each intervention, $T_1$ and $T_2$, separately. These are 8.3 and 9.5 percentage points, respectively, with 3,000 participants per site.

The detectable impacts on primary outcomes will be policy-relevant—realistic impacts consistent with those found in studies of related interventions in the welfare-to-work evaluation literature (Decker 2011; Fraker et al 2002; Bloom and Michalopoulos 2001). Based on these studies, the current study should be able to detect impacts of 3 to 6 percentage points in the

employment rate. With a sample of 5,000 per site, our expected MDIs are at the lower end of this range.

Another important standard for benchmarking the MDI is whether the study has sufficient power to detect a cost-beneficial impact from society's perspective. Assuming that $165 million is allocated evenly across 10 sites and 2,500 people receive services on average in each site, the average cost per participant would be $6,600. Assuming program benefits arise largely as a result of earnings increases, over a three-year evaluation period, quarterly earnings would have to increase by about $550 per quarter ($6,600 divided by 12 quarters) for the intervention to be cost-beneficial and less if the impacts are sustained for a longer period of time.  With a sample of 5,000, the evaluation team would be able detect this cost-beneficial impact.

Exhibit B.2.a. Minimum detectable impacts on primary and secondary outcomes for each site

| Outcome | Individual-based RA with Sample of 5,000 Participants | | Individual-based RA with Sample of 3,000 Participants | |
|---|---|---|---|---|
| | T vs. C | $T_1$ vs. C | T vs. C | $T_1$ vs. C |
| **Primary Outcomes** | | | | |
| Employment (%) | | | | |
|   Full Sample | 2.9 | 3.3 | 3.7 | 4.3 |
|   50% Subgroup | 4.0 | 4.7 | 5.2 | 6.0 |
| Quarterly earnings ($) | | | | |
|   Full Sample | 489 | 565 | 632 | 730 |
|   50% Subgroup | 692 | 799 | 894 | 1033 |
| SNAP (%) | | | | |
|   Full Sample | 2.5 | 2.9 | 3.2 | 3.7 |
|   50% Subgroup | 3.5 | 4.1 | 4.6 | 5.3 |
| **Secondary Outcomes** | | | | |
| Food security (%) at 12-month follow-up | | | | |
|   Total number of completes | 6.4 | 7.4 | 8.3 | 9.5 |
|   50% Subgroup | 10.2 | 11.7 | 13.1 | 15.1 |
| Food security (%) at 36-month follow-up | | | | |
|   Total number of completes | 8.3 | 9.5 | 10.7 | 12.4 |
|   50% Subgroup | 13.1 | 15.1 | 17.0 | 19.6 |

Notes:    Samples of individuals are allocated equally across T1, T2, and C. MDIs for comparison of T2 vs. C are the same as for T1 vs. C and are not shown in the table. Variance estimates for primary outcome measures were estimated using SIPP data from 2010 on SNAP participants ages 18 to 49 that were either employed or unemployed in the past quarter. The power analysis assumes an employment rate of 75.8%, a standard deviation of quarterly earnings of $7,333, and percentage receiving SNAP of 83.1%. The power analysis assumes that covariates in the regression model will explain 20% of the variation in the outcome measures. Variance estimates for food security was estimated using data from the SNAP-Food Security study (Mabli et al. 2013) and is based on a 34.2% food security rate. The power analysis assumes there will be a design effect due to weighting from the two-phase design of about 1.5. The total number of completes at the 12-month follow-up and 36-month follow-up are 1,840 and 1,110 for sites with 5,000 participants, respectively. The total number of completes at the 12-month follow-up and 36-month follow-up are 1,104 and 666 for sites with 3,000 participants, respectively.

Finally, the evaluation team will analyze enhancements to site interventions to learn about key program features that have national implications. The analysis will examine the effects of key program features that have national implications by pooling observations across sites with similar features, such as low- and high-intensity case management. FNS assumes that pooling

will be possible for 5 of the 10 pilot projects, resulting in smaller MDIs than presented in Table B.2a.

## 3.  Statistical methodology for sample section

The evaluation team will select samples of individuals to collect 12-month and 36-month follow-up data. Within each pilot site the evaluation team will randomly select individuals using a stratified random design (equal probability selection within strata). Strata will be formed based on research group and, if applicable, pilot location. For example, for the California pilot, 27 strata will be formed based on research groups (T, C1 and C2) and also based on the nine locations in Fresno County where the pilot will be implemented. For the Vermont pilot which will be implemented statewide, strata will be formed based only on research groups (treatment and control). For each grantee, the evaluation team will select a random sample of study participants within each stratum using equal probability sampling and will implicitly stratify individuals within strata using specific items from the registration document such as gender and age. The evaluation team does not foresee any unusual problems requiring specialized sampling procedures beyond the stratified methods proposed above.

Although the evaluation team would like to interview the random subsamples of research sample members annually to provide information on secondary outcomes such as food security, housing status, and well-being, the evaluation team have opted for follow-up of two strategically selected points of time, at 12-months and 36-months after random assignment, to assess impacts on short-term and longer-term impacts, while at the same time minimizing burden on respondents.

## B.3. Methods to maximize response rates and deal with nonresponse

**Describe methods to maximize response rates and to deal with issues of non-response.**

**The accuracy and reliability of information collected must be shown to be adequate for**

**intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.**

FNS anticipates that all Grantees (State Agencies and Business-for-and-not-for-profit) involved will participate fully in this data collection.

The SNAP E&T Pilots evaluation study team will employ a variety of methods to maximize response rates and deal with nonresponses from the individuals/household SNAP participants. This section describes these methods for the survey and for focus group data collection efforts.

## 1. Survey Data Collection

To maximize response rates for the follow-up surveys, the evaluation team will (1) have regular contact with respondents over the study period; (2) offer monetary and non-monetary incentives; (3) employ intensive locating efforts; (4) create a well-designed instrument, and (5) conduct two-phase sampling. Each of these is discussed in detail below.

The evaluation team will keep in touch with respondents throughout the study period and encourage them to notify the team about any changes in their contact information via a toll-free project telephone number. In addition to a welcome packet sent after RA that includes a study brochure (Attachment F.1 and F.2), the evaluation team will send survey reminder postcards (Attachment R.1 and R.2) throughout the field period to remind respondents of the study.

Monetary and non-monetary incentives will also be used. Research has long shown that monetary incentives provided to survey respondents can improve response rates without affecting data quality (Singer and Kulka 2002; Singer et al. 1999). For the 12-month follow-up survey (Attachments O.1 and O.2), the evaluation team offered $30 to all households for the 12-month follow-up survey (which will close December 2018) and will offer $40 for the 36-month follow-up survey. The survey incentives proposed for the SNAP E&T Pilots Evaluation are

based on the characteristics of the study population and experience with conducting telephone

surveys with similar low-income populations:

- The USDA-sponsored Supplemental Nutrition Assistance Program on Food Security (OMB Control Number 0584-0563, Discontinued September 19, 2011) offered a modest $2 pre-pay incentive and a $20 post-pay upon completing the telephone interview and had a response rate of 56 percent for baseline and 67 percent for a six-month follow-up, indicating $20-22 is not sufficient for helping to achieve the target response rate of 80 percent in SNAP E&T Pilots.

- Site-specific baseline survey response rates in the USDA-sponsored 2012 SEBTC study (OMB Control Number 0584-0559, Discontinued March 31, 2014) ranged from 39 percent to 79 percent across 14 sites using a $25 incentive. The average unweighted response rate was 67 percent; the rate was 53 percent for passive consent sites and 75 percent for active consent sites (Briefel et al. 2013), indicating that $25 was not sufficient for helping to achieve the target response rate for the proposed data collection.  Despite not achieving the target response rate, the increase in incentive for the 2012 full demonstration year from $10 to $25 proved effective in improving the response rate by 12% unweighted and 15%  weighted over the 2011 pilot year.  The increased incentive for the 2012 survey was successful in addressing respondent fatigue that was evident during the 2011 pilot year.

 Both of these recent studies conducted with populations similar to SNAP E&T Pilots

indicate that a $22-25 incentive is insufficient in reaching the higher end of the target response

rate of 80 percent. As such, the evaluation team proposed a $30 proposed incentive amount for

the 12-month survey and a $40 incentive for the 36-month survey.  Increasing the amount of the

incentive to $40 for the 36-month survey should help keep more respondents engaged, reduce

respondent fatigue with the burden involved in the number of data collection activities required

over time, and minimize response bias in the study.

 Mercer et al. (2015) conducted a meta-analysis of the dose-response association between

incentives and response and found a positive relationship between higher incentives and response

rates for household telephone surveys offering post-pay incentives. Singer et al. (1999) had in a

previous meta-analysis found that incentives in face-to-face and telephone surveys were effective

at increasing response rates, with a one dollar increase in incentive resulting in approximately a

one-third of a percentage point increase in response rate, on average.  Further, sufficient

incentives can help obtain a high cooperation rate for both the baseline and follow-up surveys so that at follow-up less field interviewer effort will be needed to locate sample members to complete the survey.

The above discussion summarizes evidence for the effectiveness of incentives for reducing non-response bias, and the response rates associated with offering lower incentive amounts to highly similar target populations. In addition, the justifications for offering incentives and the amounts to be offered are justified for several reasons that address key Office of Management and Budget (OMB) considerations (Graham 2006):

- **Improved data quality.** Incentives can increase sample representativeness. Because they may be more salient to some sample members than others, respondents who would otherwise not consider participating in the surveys may do so because of the incentive offer (Groves et al. 2000).

- **Improved coverage of specialized respondents.** Some of the populations targeted by the Pilot programs include homeless and ex-offenders, which are considered hard-to-reach (Bonevski et al., 2014). In addition, households in some of the pilot areas are specialized respondents because they are limited in number and difficult to recruit, and their lack of participation jeopardizes the impact study. Incentives may encourage greater participation among these groups.

- **Reduced respondent burden.** As described above, the incentive amounts planned for the SNAP E&T Pilots Evaluation are justified because they are commensurate with the costs of participation, which can include cellular telephone usage or travel to a location with telephone service, particularly for the homeless population served by some of the pilot programs.

- **Complex study design.** The participant surveys collected for the impact study are longitudinal. Participants will be asked to complete a registration document and two surveys over a period of 36 months. Incentives in amounts similar to those planned for this evaluation have been shown to increase response rates, decrease refusals and noncontacts, and increase data quality compared to a no-incentive control group in a longitudinal study (Singer and Ye 2013).

- **Past experience.** The studies described above demonstrate the effectiveness of incentives for surveys fielded to similar low-income study populations.

- **Equity.** The incentive amounts will be offered equally to all potential survey participants. The incentives will not be targeted to specific subgroups, nor will they be used to convert refusals. Moreover, if incentives were to be offered only to the most disadvantaged individuals, such as the homeless or ex-offenders, the differing motivations to participate

used across projects will limit the ability to compare results across target populations and sites.

In summary, the planned incentives for the longitudinal household surveys are designed to promote cooperation and high data quality and to reduce participant burden and participant costs associated with the surveys, which are similar in length and will be conducted with similar populations as in other OMB-approved information collections. If all of the other strategies to achieve high response rates are used without the planned incentives, the non-response bias will be higher, resulting in poor data quality. The likelihood of detecting project impacts will be significantly compromised.

The plan for intensive locating includes both in-house and in-field locating. Before follow-up data collection begins, the evaluation team will send a file that includes available contact information for the subsample of respondents selected for the follow-up to commercial vendors, such as Accurint, to verify and update contact information. For sample members without valid phone numbers, in-house locators will use a variety of resources, for example directory assistance and major national databases—including the U.S. Postal Service National Change of Address database; State Department of Motor Vehicles records; and death indices—to track down contact information. When necessary, locating staff will reach out to family members, friends, and neighbors identified by respondents in the registration document. This in-house location effort will reduce the number of cases that require field location, which has an impact on the project schedule and ultimately on response rates.

In-field locating will be used to find respondents that cannot be located through in-house locating efforts as well as respondents who have not been reached by telephone. When respondents are located, they will be asked to complete the survey by calling in to FNS's contractor Survey Operations Center with a cell phone that field staff will provide. The field staff

will wait as the respondent completes the telephone interview.  Once completed, the field staff

will give the respondent the gift card, and retrieve the cell phone. This approach—telephone

interviewing with in-house location followed by field follow-up—was used successfully on the

Summer Electronic Benefits Transfer for Children (SEBTC) Demonstration (OMB Number:

0584-0512, Expiration Date: 09/30/2012) which achieved response rates of 73 percent in the

spring and 80 percent in the summer, even with shorter field periods than the field period for this

evaluation.

To further maximize response rates, the evaluation team developed minimally burdensome

follow-up surveys (Attachments O.1, O.2, O.4, and O.5) that take no more than 32 minutes to

complete. To reduce respondent burden, information that can be accessed through administrative

data will not be asked for in the survey. In addition, all instruments will be available in English

and Spanish. For CATI surveys, bilingual interviewers can conduct the interview in Spanish and

possibly other languages, if deemed necessary.

To address whether the pilots significantly affect secondary outcomes, such as food security,

well-being, and housing status, the evaluation team will employ a two-phase sampling approach

that has been implemented successfully in large national surveys, such as the American

Community Survey and the National Survey of Family Growth. Two-phase sampling or double

sampling is an efficient use of resources and involves a phase 1 sampling from the full

population and then, in phase 2, selecting a second sample from the nonrespondents from phase

1. In the second phase, more focused resources (in this case, field follow-up) is devoted to the

sample, which leads to higher response rates among the phase 1 nonrespondents than what would

happen in the absence of the more focused field follow-up. This sampling approach typically

increases response rates, reduces the potential for nonresponse biases, and reduces costs and

burden associated with survey data collection, while maintaining a sufficient level of power to detect differences in outcomes between treatment and control groups within pilot sites.

The evaluation team has developed and refined methods to build rapport and overcome the reluctance of sample members to participate in the interview. Our contractor is trained in multi-pronged approaches to focus on preventing and converting refusals. The strategies aim to convince sample members that (1) the study is legitimate and worthwhile, (2) their participation is important and appreciated, and (3) the information provided will be held private and will not affect their job or their eligibility for SNAP or other benefits.

The multi-pronged approach to completing interviews includes flagging telephone refusals in the CATI scheduler and, one week after a refusal, sending a survey refusal letter (Attachment S.1 and S.2) to the sample member emphasizing the importance of the study and addressing typical concerns. Interviewers skilled at converting refusals will then contact the sample member by telephone; these trained interviewers are well prepared to address common respondent concerns that may lead to refusals. Data collectors will be selected if they have solid experience and performance in comparable topic areas, and who have demonstrated (1) reliability, (2) communication skills, (3) accurate reading and recording proficiency, and (4) aptitude for the administrative and reporting requirements of survey work.

To further encourage cooperation, the evaluation team will use certified bilingual interviewers to complete interviews in Spanish. Moreover, to develop the skills necessary to encourage participation among low-income households, all telephone interviewers receive general interviewer training before being assigned to a study. This training involves essential interviewing skills, probing, establishing rapport, avoiding refusals, eliminating bias, and being sensitive to at-risk and special populations. In addition, all interviewers will receive project-

specific training that reviews study goals, provides a question-by-question review of the

instrument, and conveys best practices for interviewing for the specific study. Interviewers will

also conduct mock interviews prior to conducting real interviews. Data collector training

sessions will take place no earlier than two weeks before data collection begins.

## 2. Focus group data collection

Once individuals were recruited (see Part A, section A.2, 3d) to participate in focus groups,

a personally addressed focus group confirmation letter (Attachment L) was prepared and sent to

each individual who agreed to participate confirming the time, date, and place of the focus group.

The materials also noted the $50 incentive for participating in the focus group.

B.4. Description of tests of procedures

**Describe any tests of procedures or methods to be undertaken. Testing is encouraged
as an effective means of refining collections of information to minimize burden and
improve utility. Tests must be approved if they call for answers to identical questions
from 10 or more respondents. A proposed test or set of tests may be submitted for
approval separately or in combination with the main collection of information.**

To develop the follow-up surveys the evaluation team drew on content in other surveys that

have been administered successfully with similar low-income populations, such as the surveys

used in the Workforce Investment Act Gold-Standard Evaluation (WIA GSE) or Rural Welfare-

to-Work, and then pretested the survey instruments. Since most survey items were drawn from

previously administered surveys which have been tested and used successfully, the pretest

focused mainly on question sequences and survey processes. The survey team pretested the 12-

month follow-up survey (Attachment O.1) with nine current SNAP E&T participants from two

pilot sites. The survey instruments were refined based on pretest results. A memo describing the

key results from the pretest is included in Attachment N.

B.5. Individuals consulted on statistical aspects of the design

**Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), awardee(s), or other person(s) who will actually collect and/or analyze the information for the agency.**

The information will be collected and analyzed by Mathematica Policy Research. The sampling procedures were developed by Nicholas Beyler (telephone 202-250-3539) of Mathematica. The sampling plans were reviewed internally by Michael Sinclair (telephone 202-552-6439), a senior fellow at Mathematica. Jennifer Rhorer, (800) 727-9540, of the National Agricultural Statistics Service (NASS) has also reviewed this supporting statement and provided comments that have been incorporated.

## REFERENCES

Bloom, D., and C. Michalopoulos. "How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research." New York: MDRC, 2001.

Decker, Paul T., and Jillian Berk. "Ten Years of the Workforce Investment Act (WIA): Interpreting the Research on WIA and Related Programs." *Journal of Policy Analysis and Management,* vol. 30, no. 4, fall 2011, pp. 906–926.

Fraker, T.M., C.M. Ross, R.A. Stapulonis, R.B. Olsen, M.D. Kovac, M.R. Dion, and A. Rangarajan. "The Evaluation of Welfare Reform in Iowa: Final Impact Report." Washington, DC*:* Mathematica Policy Research, 2001.

Mercer, A., A. Caporaso, D. Cantor, and R. Townsend, R. "How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys." *Public Opinion Quarterly,* 2015, vol. 79 (1), pp.105-129.

Singer, E., and R.A. Kulka. "Paying Respondents for Survey Participation." In *Studies of Welfare Populations: Data Collection and Research Issues. Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs,* edited by Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press, 2002, pp. 105–128.

Singer, E., R.M. Groves, and A.D. Corning. "Differential Incentives: Beliefs About Practices, Perceptions of Equity, and Effects on Survey Participation." *Public Opinion Quarterly,* vol. 63, 1999, pp. 251–260.